



European Soccer Database

(Regression)

Bernardo Costa Moreira

up201604014

Filipe Carlos de Almeida Duarte da Cunha Nogueira

up201604129

Francisco Jorge de Almeida Henriques Pereira

up201605306



(1) especificação do trabalho a realizar (definição do(s) problema(s) de aprendizagem computacional a resolver);

Para este projeto pretende-se aprender a prever valores para uma determinada partida de futebol, dada uma série de informação relevante.

Inicialmente é nos fornecida uma base de dados, que terá de ser pré-processada e eventualmente será necessário efetuar uma transformação e avaliação dos dados.

Com os dados devidamente tratados pretendemos encontrar padrões que permita ao nosso programa chegar ao valor pretendido e utilizar diferentes algoritmos para comparar o desempenho dos mesmos.



(1) especificação do trabalho a realizar (definição do(s) problema(s) de aprendizagem computacional a resolver);

Dado que o modelo deve incidir num problema de regressão, vamos treinar o mesmo para que este faça uma previsão quanto ao **número de golos marcados**. Para isso definimos como parâmetros de entrada:

- Capacidade ofensiva / Capacidade do meio-campo/ Capacidade defensiva da equipa a jogar a partida
- Classificação média dos jogadores a jogar essa mesma partida
- Altura / Peso médio dos jogadores a jogar essa mesma partida
- Média de ods para a partida (Casa/Empate/Fora)
- Número de vitórias seguidas
- Número de derrotas seguidas
- Índice de Vitórias



(2) trabalho relacionado com referências a trabalhos encontrados na pesquisa (artigos, páginas web e/ou código fonte);

- <https://towardsdatascience.com/the-5-feature-selection-algorithms-every-data-scientist-need-to-know-3a6b566efd2>
- <https://medium.com/data-hackers/implementando-regress%C3%A3o-linear-simples-em-python-91df53b920a8>
- <https://elitedatascience.com/python-machine-learning-tutorial-scikit-learn#step-1>
- <https://www.imperial.ac.uk/media/imperial-college/faculty-of-engineering/computing/public/1718-ug-projects/Corentin-Herbinet-Using-Machine-Learning-techniques-to-predict-the-outcome-of-professional-football-matches.pdf>
- https://www.researchgate.net/publication/257048220_Machine_Learning_for_Soccer_Analytics



(3) descrição das ferramentas e algoritmos a utilizar no trabalho;

Ferramentas a utilizar durante o trabalho:

- RapidMiner - visualização e análise de dados;
- Jupyter Notebooks - desenvolvimento de código;
- Scikit-learn & Pandas- ferramentas de auxílio ao pré-processamento de dados e utilização de algoritmos de machine learning.



(3) descrição das ferramentas e algoritmos a utilizar no trabalho;

Para testar diferentes desempenhos do modelo pretendemos integrar no mesmo 3 diferentes algoritmos:

- Árvores de Decisão;
- Redes Neurais;
- Máquinas de Suporte Vetorial;



(4) trabalho de implementação já realizado.

Atualmente todo o trabalho realizado foi dispendido na análise e tratamento de dados:

- Os valores nulos foram removidos, a base de dados foi dividida por tabelas e algumas colunas não relevantes das tabelas foram removidas.
- Com estas alterações realizadas, construímos os atributos de entrada relativos à própria partida, por forma a facilitar a análise e a aprendizagem do algoritmo.